

# Curso Inteligência Artificial: do Zero ao Infinito

## Overfitting e Métodos de Regularização

Universidade Federal de Mato Grosso

# Agenda

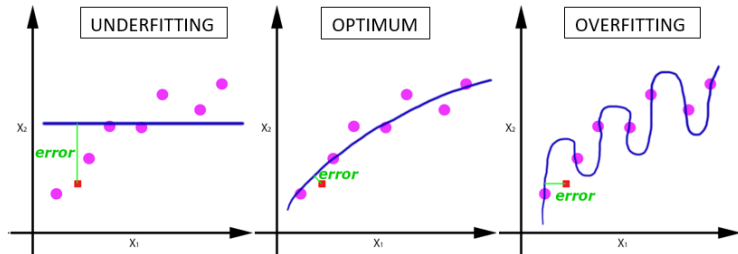
- 1 Introdução
- 2 Keep it Simple
- 3 Early stopping
- 4 Dropout
- 5 Regularização L1 e L2

# Agenda

- 1 Introdução
- 2 Keep it Simple
- 3 Early stopping
- 4 Dropout
- 5 Regularização L1 e L2

- **Overfitting** é um dos principais problemas durante o treinamento de modelos neurais.
- Ocorre quando um modelo realiza sua tarefa de forma excelente no conjunto de treino, mas é ineficaz em um conjunto de teste.
- Evitar o *overfitting* pode, por si só, melhorar o desempenho do modelo.

# Overfitting



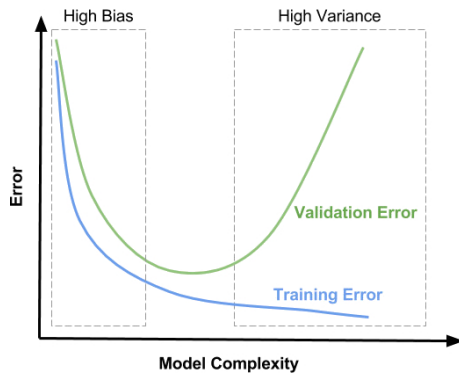
# Introdução

- Variância indica a variabilidade das previsões do modelo.
- Um modelo com *alta variância* “decorou” o conjunto de treinamento, mas não consegue prever dados não vistos.
- Tornou-se muito complexo de tal forma que não é capaz de generalizar resultados (**overfitting**).

# Introdução

- *Viés* é a diferença entre as previsões do modelo e os valores reais.
- Ocorrem altas taxas de erro no conjunto de treinamento e também no conjunto de testes (**underfitting**).
- Um modelo demasiadamente simplificado ou mal treinado resulta em uma previsões falhas e pouco assertivas.

# Complexidade do Modelo





# Redes Neurais Multicamadas

## Referências

- Understanding the Bias-Variance Tradeoff

# Agenda

- 1 Introdução
- 2 Keep it Simple**
- 3 Early stopping
- 4 Dropout
- 5 Regularização L1 e L2

# Keep it Simple

- **Simplificar a arquitetura** é o primeiro passo para evitar *overfitting*.
- Remover algumas camadas.
- Diminuir número de neurônios.
- Infelizmente, não existe uma regra universal.

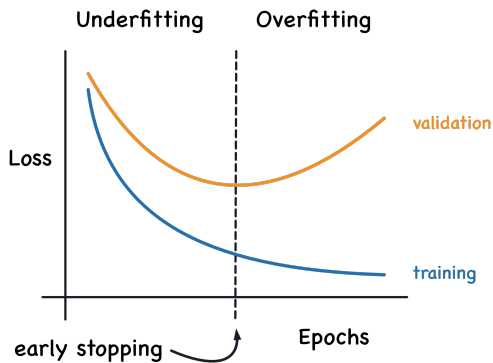
# Agenda

- 1 Introdução
- 2 Keep it Simple
- 3 Early stopping**
- 4 Dropout
- 5 Regularização L1 e L2

# Early stopping

- **Early stopping** consiste em encerrar o treinamento antes que aconteça *overfitting*.
- Interrompe-se o treinamento quando ocorre divergência entre os valores do custo de treinamento e de teste.

# Early stopping



Referência: <https://www.kaggle.com/ryanholbrook/overfitting-and-underfitting>

# Referências

## Early stopping

- A Gentle Introduction to Early Stopping to Avoid Overtraining Neural Networks
  - ▶ <https://machinelearningmastery.com/early-stopping-to-avoid-overtraining-neural-network-models/>
- Introduction to Early Stopping: an effective tool to regularize neural nets
  - ▶ <https://towardsdatascience.com/early-stopping-a-cool-strategy-to-regularize-neural-networks-bfdeca6d722e>

# Agenda

- 1 Introdução
- 2 Keep it Simple
- 3 Early stopping
- 4 Dropout**
- 5 Regularização L1 e L2

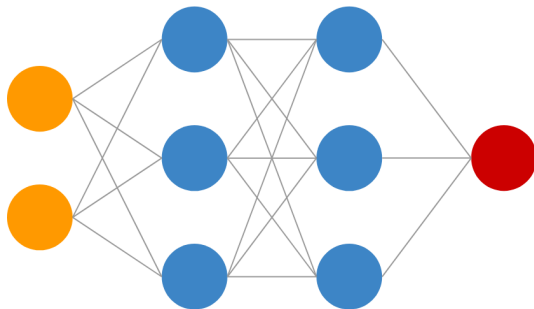


# Dropout

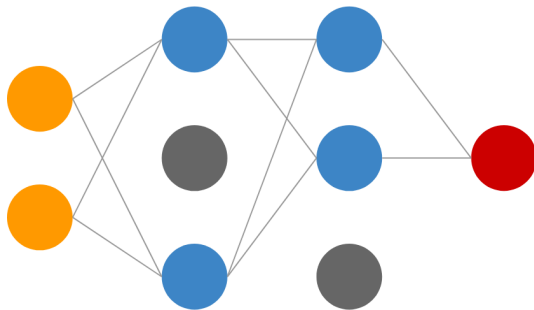
- **Dropout** consiste em aleatoriamente “matar” alguns neurônios da rede neural durante o treinamento.
- Previne que os neurônios se especializem demais em algumas instâncias do problema.
- Torna o modelo mais robusto, fazendo com que alguns neurônios aprendam a corrigir erros dos demais.

Referência: Dropout: A Simple Way to Prevent Neural Networks from Overfitting

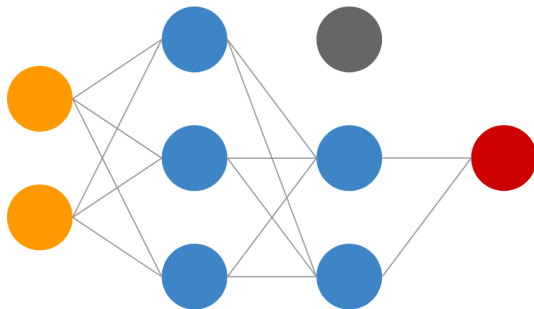
# Dropout



# Dropout



# Dropout



# Dropout

- Define-se uma probabilidade de neurônios que sofrerão *dropout*.
  - ▶  $0,1 \leq p \leq 0,5$
- *Dropout* deve ser aplicado nas camadas ocultas.
- Não deve-se aplicar na camada de saída.

## Origem do termo

*The term “dropout” refers to dropping out units (hidden and visible) in a neural network.*

Referência: A Gentle Introduction to Dropout for Regularizing Deep Neural Networks

# Referências

## Dropout

- Dropout: A Simple Way to Prevent Neural Networks from Overfitting
  - ▶ <https://jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf>
- Deep Learning Book: Capítulo 23 - Como Funciona o Dropout?
  - ▶ <https://www.deeplearningbook.com.br/capitulo-23-como-funciona-o-dropout/>
- A Gentle Introduction to Dropout for Regularizing Deep Neural Networks
  - ▶ <https://machinelearningmastery.com/dropout-for-regularizing-deep-neural-networks/>

# Agenda

- 1 Introdução
- 2 Keep it Simple
- 3 Early stopping
- 4 Dropout
- 5 Regularização L1 e L2**

# Regularização L1 e L2

- Regularização L1 e L2 adicionam uma penalidade à função erro.
- A ideia é que ao invés de “zerar” alguns pesos, podemos simplesmente reduzir seu valor.
- Minimizar a norma incentiva os pesos a serem pequenos, o que, por sua vez, fornece funções “mais simples”.



# Regularização L1 e L2

L1 e L2 referem-se à Norma L1 e L2

$$\text{L1 Norm } ||w||_1 = \sum_i^n |w_i|$$

$$\text{L2 Norm } ||w||_2 = \sum_i^n w_i^2$$

# Regularização L1

- A norma L1 levará alguns pesos a 0, induzindo esparsidade nos pesos.
- Os pesos diminuem em uma quantidade constante para 0.
- Isso evita que a rede “decore” algumas instâncias.

$$E(p, y) = \frac{1}{2}(p - y)^2 + \lambda \sum |w_i|$$

Referência: <https://ichi.pro/pt/visualizando-a-regularizacao-e-as-normas-l1-e-l2-239017789921586>

# Regularização L2

- A norma L2 reduzirá todos os pesos, mas não totalmente para 0.
- Os pesos diminuem em um valor proporcional a  $w$ .
- É menos eficiente em relação à L1 quanto à memória neural.

$$E(p, y) = \frac{1}{2}(p - y)^2 + \lambda \sum |w_i|^2$$

Referência: <https://ichi.pro/pt/visualizando-a-regularizacao-e-as-normas-l1-e-l2-239017789921586>

# Exemplo Prático MNIST

Referência: <https://selectstar-ai.medium.com/what-is-mnist-and-why-is-it-important-e9a269edbad5>

# Referências

## Regularização L1 e L2

- Intuitions on L1 and L2 Regularization
  - ▶ <https://towardsdatascience.com/intuitions-on-l1-and-l2-regularisation-235f2db4c261>
- Visualizando a regularização e as normas L1 e L2
  - ▶ <https://ichi.pro/pt/visualizando-a-regularizacao-e-as-normas-l1-e-l2-239017789921586>
- Deep Learning Book: Capítulo 22 - Regularização L1
  - ▶ <https://www.deeplearningbook.com.br/capitulo-22-regularizacao-l1/>

# Curso Inteligência Artificial: do Zero ao Infinito

## Overfitting e Métodos de Regularização

Universidade Federal de Mato Grosso